

Elementi di Psicometria con Laboratorio di SPSS 1

19-Rapporto fra varianze e 20-Introduzione all'Anova
vers. 1.0 (5 dicembre 2014)
versione per stampa

Germano Rossi¹

`germano.rossi@unimib.it`

¹Dipartimento di Psicologia, Università di Milano-Bicocca

2014-2015

Saltare

dal penultimo paragrafo di p.203 fino alla fine del capitolo.

Inizio del paragrafo “Non abbiamo informazioni certe...”.

Rapporto fra varianze

- Ci sono diverse circostanze in cui c'è bisogno di lavorare con le varianze al posto delle medie
- Il t-test standard necessita che le varianze dei due gruppi da confrontare sia uguale (ma non sempre è così)
- Se i due gruppi hanno varianza diversa, si usa una formula di t che tiene in considerazione la diversità delle varianze
- Se si hanno più di due gruppi il t-test non può essere usato (salvo rischiare di “sforare” l'errore di I tipo)
- Per questo motivo esiste un test per verificare il **rapporto di due varianze** e una tecnica per studiare più gruppi in contemporanea

Rapporto fra varianze

- Il rapporto di due varianze si distribuisce secondo la curva di F (di Snedecor che l'ha studiata)
- F è anche la sigla della statistica che indica il rapporto di varianze
- La formula è semplicemente

$$F = \frac{\text{varianza maggiore}}{\text{varianza minore}}$$

- le varianze utilizzate sono le *stime* a partire dal campione
- F è sempre a una coda ed è (dovrebbe essere) ≥ 1

Rapporto fra varianze

- Essendo un rapporto di 2 varianze, esistono 2 gradi di libertà
- e le tavole di F sono sempre complicate e lunghe
- Qui c'è l'inizio della tavola 19.2 (p.197) per $\alpha = .05$

Gradi di libertà relativi alla stima della varianza minore (denominatore)	Gradi di libertà relativi alla stima della varianza maggiore (numeratore)					
	5	7	10	20	50	∞
5	5.1 o maggiore	4.9	4.7	4.6	4.4	4.4
6	4.4	4.1	4.1	3.9	3.8	3.7
7	4.0	3.8	3.6	3.4	3.3	3.2

Confronti fra più gruppi

- Se devo confrontare 3 medie?
- Faccio 3 confronti: M_1 con M_2 , poi M_1 con M_3 e quindi M_2 con M_3
- Equivale a una una combinazione di 3 elementi presi 2 a 2 senza ripetizione.
- Con 4 medie?
- Faccio 6 confronti!
- Con 5 medie?
- Faccio 10 confronti
- Ma...

Più confronti di medie

- Se per ogni confronto, ho una possibilità di sbagliare pari al livello α che scelgo, per 10 confronti avrò una possibilità pari a 10 volte α
- Ovvero, se $\alpha = .05$, $10 \times 0.05 = 0.50$
- Ciò significherebbe che su 10 confronti almeno la metà potrebbero essere inaffidabili
- È chiaro che non posso correre un rischio così elevato.
- Ho due possibili soluzioni
 - Analisi della varianza (preferibile)
 - Criterio di Bonferroni: dividere l'alfa per il numero di confronti e usare il risultato come nuovo alfa (es. $\alpha = .05$ con 10 confronti, $\alpha/10 \Rightarrow .05/10 = .005$)

Analisi della varianza

- Il test dell'*Analisi della varianza*, risolve questo problema
- anche conosciuto con i nomi Anova o AOV (da *Analysis of variance*)
- Serve per confrontare fra loro tre o più gruppi e decidere se vengono dalla stessa popolazione di riferimento
- Affronteremo l'anova a livello semplicistico

Tipi di anova

- Le situazioni più comuni sono:
 - 1 variabile dipendente (I/R) suddivisa in base ad 1 variabile indipendente che ha 3 o più categorie (N/O): *Anova ad 1 via (o 1 fattore)*
 - 1 variabile dipendente (I/R) suddivisa in base a 2 o più variabili indipendenti (N/O): *Anova a 2 o più vie (o più fattori)*
 - 1 variabile dipendente misurata più volte (*Anova a misure ripetute*)
 - 1 variabile dipendente e 1 o più variabili indipendenti (*Anova mista*)

Tipi di anova

- I modelli anova si distinguono anche come:
 - modelli *between, tra*: tutte le osservazioni sono indipendenti rispetto alle variabili indipendenti (estensione del t-test per campioni indipendenti)
 - modelli *within, entro*: le variabili dipendenti misurano più volte lo stesso caso statistico (estensione del t-test appaiato)
 - modelli *misti*: ci sono sia variabili misurate più volte sia variabili dipendenti
- Per spiegare i concetti userò l'anova a fattori indipendenti e soltanto tramite SPSS

Concetto intuitivo

$$\frac{\begin{array}{|c|c|c|} \hline & & \\ \hline \end{array}}{\begin{array}{|c|} \hline \\ \hline \end{array}} = F$$

- Con sintesi semplicistica, possiamo dire che l'AOV confronta la varianza calcolata in due modi diversi:
 - la varianza tra i singoli gruppi (ogni gruppo è considerato separato)
- la varianza entro tutti i gruppi (un solo gruppo totale ottenuto ignorando i singoli gruppi)
- la statistica che risulta è il rapporto fra due varianze:
- la media delle varianze dei singoli gruppi
- la varianza totale

Concetto intuitivo

- La statistica dell'anova è indicata con F
- Se le due stime di varianza sono uguali, F si avvicinerà ad 1
- Se le due stime di varianza sono diverse, F sarà tanto più grande quanto maggiore è la diversità
- I gradi di libertà sono 2 (uno per ogni varianza utilizzata) e dipendono dai gruppi confrontati ($k - 1$) e dalle loro numerosità (in genere $N_x - 1$, in base ai confronti)

Concetto intuitivo

- Esistono altri modo per pensare l'analisi della varianza
- Immaginate la misurazione di una variabile come se fosse composta da una misurazione vera e da un errore dovuto allo strumento di misurazione

$$\text{valore osservato} = \text{valore vero} + \text{errore}$$

- Oppure immaginate che una variabile Y (di per sé, senza l'influenza di altre variabili) sia una costante uguale alla media osservata
- e che ogni variabile indipendente (X, Z) produca una modifica nei risultati (in aumento o in diminuzione) come parte (a, b) di se stessa

$$Y = \mu_Y \pm aX \pm bZ + \text{errore}$$

Per capire meglio

- se F è un rapporto di varianze, dobbiamo considerare meglio la varianza:

$$\text{var}_{stimata} = \frac{\sum (X - \bar{X})^2}{N - 1} = \frac{\text{somma dei quadrati}}{\text{gradi di libertà}}$$

- L'anova quindi fa diverse stime di somme di quadrati (numeratore)
- le divide per i gradi di libertà (denominatore)
- quindi divide fra loro due diverse stime di varianza

Esempio SPSS: anova 1 via

Orientamento politico	Scala di ortodossia RFS		
	N	Mean	Std. Deviation
Nessuno	79	5,63	8,11
Sinistra	120	-1,26	9,34
Centro	104	6,38	8,21
Destra	34	5,24	10,14
Total	337	3,37	9,43

$$4.032,535 / 3 = 1.344,178$$

$$4.032,535 / 3 = 1.344,178$$

$$1.344,178 / 77,651 = 17,310$$

	SQ=Somme quadrati	df	MQ=Medie quadrati	F	Sig.
Fra	4.032,535	3	1.344,178	17,310	0,000
Errore	25.857,839	333	77,651		
Total	29.890,374	336			

SQ è il numeratore della varianza; MQ=SQ/df ed è la varianza; F è il rapporto fra le due MQ

Esempio SPSS: anova e post-hoc

- Dal momento che l'anova è statisticamente significativa (ho rifiutato H_0) almeno un gruppo è statisticamente diverso dagli altri
- per sapere qual è, usa l'analisi a *post-hoc* che confronta ciascun gruppo con tutti gli altri
- uno dei metodi a *post-hoc* crea gruppi statisticamente omogenei

	Orientamento politico	N	Subset for alpha = .05	
			1	2
Student-Newman-Keuls	Sinistra	120	-1,2583	
	Destra	34		5,2353
	Nessuno	79		5,6329
	Centro	104		6,375
	Sig.		1	0,736

Anova e t-test

- Se si fa un'anova a 1 via con una variabile indipendente che ha solo 2 gruppi
- si vede che F è il quadrato di t
- e t è la radice quadrata di F

	t	gl	Sign.
compiacenza	-,243	58	,809

$$t = -0,243 = \sqrt{0,059}$$

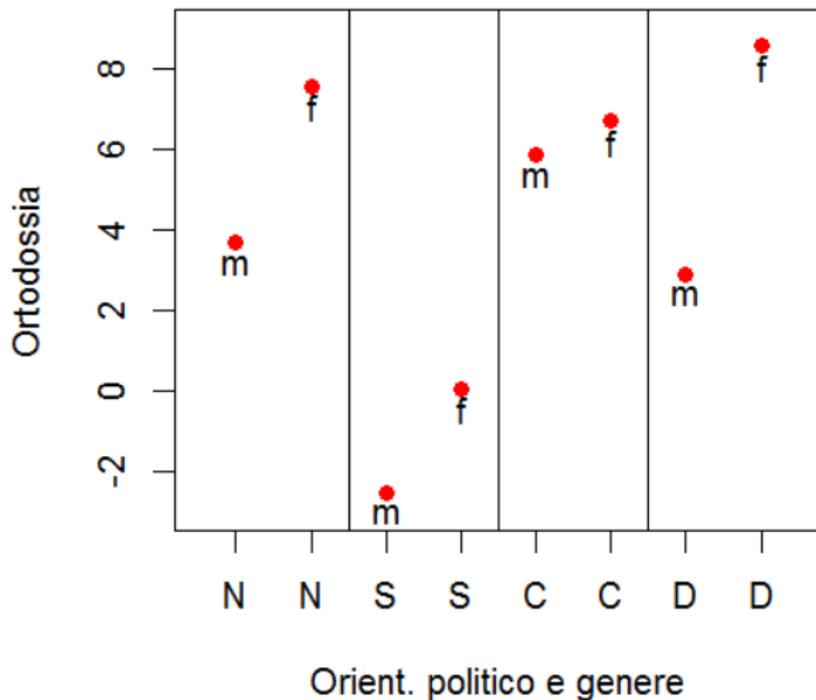
$$F = 0,059 = -0,243^2$$

	Somma quadrati	gl	Media quadratica	F	Sign.
Tra gruppi	8,929	1	8,929	,059	,809
Entro i gruppi	8787,804	58	151,514		
Totale	8796,733	59			

Esempio SPSS: Anova a 2 vie

Orientamento politico	Sesso	Scala di ortodossia		
		Mean	Std. Deviation	N
Nessuno	Maschi	3,67	9,07	39
	Femmine	7,55	6,63	40
	Total	5,63	8,11	79
Sinistra	Maschi	-2,55	9,56	60
	Femmine	0,03	9,01	60
	Total	-1,26	9,34	120
Centro	Maschi	5,85	8,62	40
	Femmine	6,70	8,00	64
	Total	6,37	8,21	104
Destra	Maschi	2,90	11,47	20
	Femmine	8,57	6,96	14
	Total	5,24	10,14	34
Total	Maschi	1,77	10,02	159
	Femmine	4,79	8,65	178
	Total	3,37	9,43	337

Esempio SPSS: Anova a 2 vie



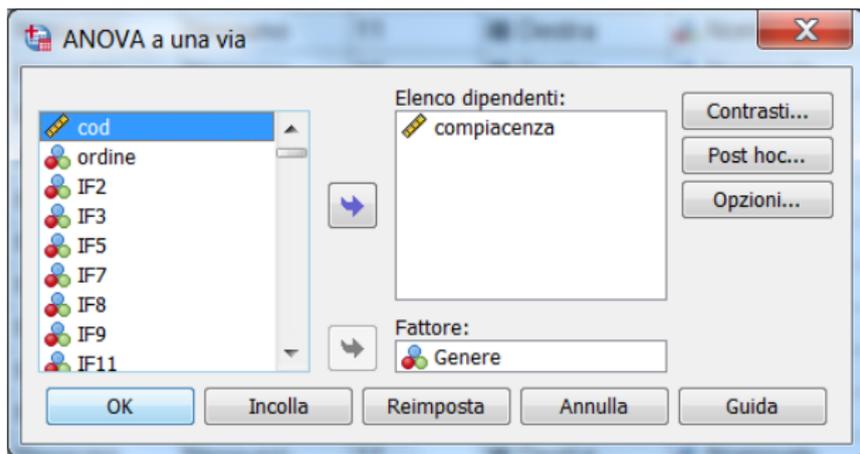
N=Nessun orientamento;
 S=Sinistra;
 C=Centro;
 D=Destra
 m=maschi;
 f=femmine

Esempio SPSS: Anova a 2 vie

Source	Type III Sum of Sq.	df	Mean Square	F	Sig.
Corrected Model	4813,336	7	687,619	9,021	0,000
Intercept	4352,743	1	4352,743	57,106	0,000
Fasce pol	3981,851	3	1327,284	17,413	0,000
sesso	685,986	1	685,986	9,000	0,003
Fasce pol * sesso	184,718	3	61,573	0,808	0,49
Error	25077,04	329	76,222		
Total	33713	337			
Corrected Total	29890,37	336			

Anova 1 via in SPSS

- Analizza | Confronta medie | Anova a 1 via
- Inserire in **Elenco dipendenti** la o le variabili quantitative
- Inserire in **Fattore** la variabile dipendente
- Infine



Anova 1 via in SPSS

- Con il bottone `Post-Hoc` si possono scegliere i post-hoc
- Generalmente si fa solo per risultati significativi

ANOVA a una via: Comparazioni multiple post hoc

Presumi varianze uguali

LSD S-N-K Waller-Duncan
 Bonferroni Tukey Rapporto di errore Tipo I/Tipo II: 100
 Sidak Tukey-b Dunnett
 Scheffé Duncan Categoria di controllo: Ultimo
 R-E-G-W F Hochberg (GT2) Test
 R-E-G-W Q Gabriel 2 lati < Controllo > Controllo

Non presumere varianze uguali

Tamhane (T2) T3 di Dunnett Games-Howell C di Dunnett

Livello di significatività: 0,05

Continua Annulla Guida

Anova 2 vie in SPSS

- Analizza | Modello lineare generalizzato | Univariata...
- Inserire in Variabile dipendente una variabili quantitative
- Inserire in Fattori fissi le variabili dipendenti
- Infine

